527, 911

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

Rec'd PCT/PTO 14 MAR 2005

(43) International Publication Date
25 March 2004 (25.03.2004)

PCT

(10) International Publication Number
WO 2004/025557 A2

(51) International Patent Classification[7]: G06T 5/00

(21) International Application Number:
PCT/GB2003/003978

(22) International Filing Date:
12 September 2003 (12.09.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0221144.9 12 September 2002 (12.09.2002) GB

(71) Applicant *(for all designated States except US)*: **SNELL & WILCOX LIMITED** [GB/GB]; 6 Old Lodge Place, St. Margaret's, Twickenham, Middlesex TW1 1RQ (GB).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **KNEE, Michael, James** [GB/GB]; 6 Woodbury Avenue, Petersfield, Hants GU32 2EE (GB). **WESTON, Martin** [GB/GB]; 7B Weston Road, Petersfield, Hampshire GU31 4JF (GB).

(74) Agents: **GARRATT, Peter, Douglas** et al.; Mathys & Squire, 100 Gray's Inn Road, London WC1X 8AL (GB).

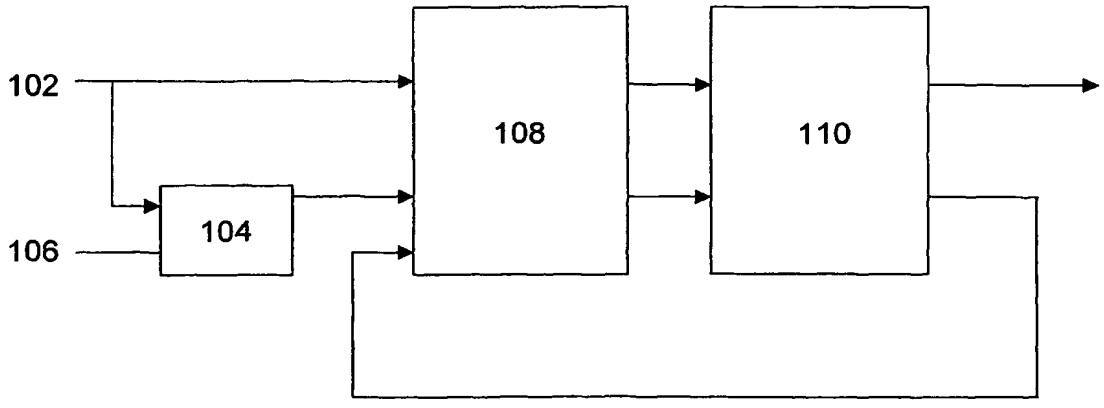
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: IMAGE PROCESSING



(57) Abstract: Video data is segmented by representing the pixel location, RGB values and other features such as motion vectors, as points in a multidimensional segmentation space. Initialized segments are represented as locations in the segmentation space and segment membership then determined by the distance in segmentation space from the data point representing the pixel to the location of the segment. The distance measure takes into consideration the covariance of the data, for the segment or for the picture.

WO 2004/025557 A2

IMAGE PROCESSING

This invention is directed to image segmentation and in an important example to the segmentation of sequences of images, such as video.

There are many methods of image segmentation in existence. Classical methods of segmenting an image described by a one-dimensional parameter (e.g. luminance) fall into two categories, those based on edge detection (e.g. the Sobel operator) and those based on region detection (e.g. sieves, watershed transforms). Work has also been done on segmentation of images described by two or more parameters (e.g. colour components). For this purpose, histogram processing is often used. It has been suggested that a model of gravitational clustering be employed in RGB measurement space, (see Yung et al "Segmentation of colour images based on the gravitational clustering concept" Optical Engineering, Soc of Photo-Optical Instrumentation Engineers 37 No.3 1 March 1998 pp 989-1000). It has also been suggested to employ gravitational clustering in a measurement space which includes not only colour information, but also location information for each pixel, (see, Hwajeong et al "Colour image segmentation based on clustering using color space distance and neighbourhood relation among pixels" Journal of Korea Inf. Sci. Soc. Software and Applications Oct. 2000, Vol. 27 No. 10 pp 1038-1045).

The prior art methods of segmentation require excessive computational processing before useful results can be achieved. The concept of gravitational clustering depends upon on a relatively large number of iterations per image, with the "mass" of each segment changing as the iterations proceed. Small changes in the initialization and termination conditions have marked and not always predictable effects on performance. Although the use of a multi-dimensional space (incorporating both colour space and pixel location) is beneficial, performance becomes critically dependent on the relative scaling in the respective dimensions. Prior art methods have not been found to work well with sequences of images - video, for example - where rapid and reliable segmentation decisions are essential.

It is an object of aspects of the present invention to provide improved methods of image segmentation that are capable of providing rapid and reliable segmentation decisions.

Accordingly, the present invention consists in one aspect in a method of segmenting image data having a plurality of feature values at each pixel, in which the data is represented as points in a segmentation vector space which is the product of the vector space of feature values and the vector space of pixel addresses, wherein segments are represented as locations in the segmentation vector space, and the membership of a segment for each pixel is determined by the distance in segmentation vector space from the data point representing the pixel to the location of the segment.

It has been found that a segmentation method in which membership of a segment is determined by an appropriate measure of distance provides more reliable segmentation than can be achieved by gravitational clustering. It has also been found that methods according to this invention, when appropriately initialized, can reach segmentation decisions without the need for large numbers of iterations. In the case - for example - of video, segment decisions from the preceding image will in many cases provide sufficiently good initialization for reliable segmentation to be achieved in a single step.

Accordingly, the present invention consists in another aspect in a method of segmenting video data having a plurality of feature values at each pixel in a sequence of pictures, in which the data is represented as points in a segmentation vector space which is the product of the vector space of feature values and the vector space of pixel addresses, and wherein segments are represented as locations in the segmentation vector space, the method comprising the steps for each picture of initially assigning pixels to segments according to the segment membership of the respective pixel in the preceding picture in the sequence; calculating the location in segmentation vector space for each initial segment utilising feature values from the current picture and determining the membership of a segment for each pixel according to the

distance in segmentation vector space from the data point representing the pixel to the location of the segment.

According to a further aspect, the present invention consists in a method of segmenting image data having a plurality of feature values at each pixel, in which the data is represented as points in a segmentation vector space which is the product of the vector space of feature values and the vector space of pixel addresses, comprising the steps of scaling the image data so as substantially to equalize the variance of the data in at least one dimension of the pixel address and each dimension of the feature value; initially assigning pixels to segments; representing each segment as a location in the segmentation vector space; and determining the membership of a segment for each pixel according to the distance the segmentation vector space from the data point representing the pixel to the location of the segment.

According to yet a further aspect, the present invention consists in a method of segmenting image data having a plurality of feature values at each pixel, comprising the steps of representing the image data as points in a segmentation vector space which is the product of the vector space of feature values and the vector space of pixel addresses in a toroidal canvas; initially assigning pixels to segments represented as locations in the segmentation vector space; and determining the membership of a segment for each pixel according to a distance measure from the data point representing the pixel to the representation of the segment.

By representing the pixel addresses on a toroidal canvas, the problem is avoided of the disappearance and reappearance of objects because of global motion in the scene.

The invention will now be described by way of example with reference to the accompanying drawing which is a block diagram illustrating a method according to the present invention.

The general aim of a segmentation method is to partition the image or picture into a number of segments, taking into account the location and value of each pixel. For moving sequences, it is additionally desirable to have a smooth transition in the segmentation from one picture to the next.

In the approach taken by this invention, the picture is represented in multidimensional space, which will here be called the universe. This is the product of the ordinary two-dimensional space of the picture, which we shall call the canvas, and the pixel space, which is the space in which the pixel or feature values themselves are defined. For example, when segmenting an RGB picture, the pixel space will have three dimensions and the universe will therefore have five dimensions.

The picture is represented as a set of points in the universe, one for each pixel. The co-ordinates of the pixel in the universe describe its position on the canvas together with its value in pixel space. Segmentation in this example is simply a partitioning of the set of pixels into a fixed number of subsets.

Referring to Figure 1, the method starts with some initial segmentation of the picture at terminal 102 into the desired number of segments. At the very start of a sequence, this initialization is performed in box 104 and might correspond to a straightforward division of the canvas into equal areas. More complex initialization procedures may take image data into account and may select dynamically the appropriate number of segments. The chosen number of segments may also be externally input at 106. Subsequently, the initial segmentation may be the segmentation of the previous picture. The processor at box 108 then calculates (as described more fully below) the centroid of each segment, according to the values of the segment's pixels in the universe. In box 110, a measure is then taken of the distance in the universe between each pixel and the centroid of each segment and each pixel is reassigned to the segment with the closest centroid.

These two steps of determining the location of each segment - by calculating the centroid - and assigning pixels to the closest segment, may be

repeated once or more using the same picture. The result is a new segmentation, which may be used as the initial segmentation for the next picture.

Performing several iterations is likely to be necessary when segmenting a single picture starting from trivial initial conditions. However, for moving sequences, there may be little to be gained from performing more than one iteration per picture.

A segmented version of the picture can be created by replacing the pixel values in each segment with the projections of the segment centroids onto the pixel space.

The example will be taken of a nine-dimensional segmentation vector space, representing the pixel address and seven feature values for each pixel, thus:

x_1, the horizontal spatial coordinate

x_2, the vertical spatial coordinate

x_3, the red component of the video signal expressed in 8-bit values

x_4, the green component of the video signal expressed in 8-bit values

x_5, the blue component of the video signal expressed in 8-bit values

x_6, a texture measure for the video, for example the standard deviation of the luminance over a 5 x 5 block centred on the current pixel

x_7, the horizontal component of a motion vector expressed in pixels per frame period

x_8, the vertical component of a motion vector expressed in pixels per frame period

x_9, the value of a displaced frame difference when a motion model for the segment is applied to the luminance data.

If $(x_1, x_2, \cdots, x_N)_k, k \in S$ are vectors in the multidimensional space belonging to segment S, then the centroid of the segment is given by

$$(\mu_1, \mu_2, \cdots, \mu_N) = \frac{1}{K_S} \sum_{k \in S} (x_1, x_2, \cdots, x_N)_k$$

where K_S is the number of points in segment S.

The Euclidean distance measure from a point $\mathbf{x} = (x_1, x_2, \cdots, x_N)$ to the segment *S* is then equal to $\sum_{n=1}^{N} (x_n - \mu_n)^2$

Then a suitably scaled distance measure might be given by

$$(x_1 - \mu_1)^2 + (x_2 - \mu_2)^2 + 0.5(x_3 - \mu_3)^2 + 1.5(x_4 - \mu_4)^2 + 0.3(x_5 - \mu_5)^2 + 3x_6^2 + 10(x_7 - \mu_7)^2 + 10(x_8 - \mu_8)^2 + 2x_9^2$$

The performance of a segmentation method depends on the relative scaling of the co-ordinate axes in the universe. For example, if on the one hand the canvas co-ordinates are multiplied by a large scaling factor, the algorithm will tend to pay more attention to the spatial component of the overall Euclidean distance and the final segmentation will look more like a simple partitioning of the canvas. If on the other hand the pixel space co-ordinates are given a large scaling factor, the segmentation will be dominated by pixel values and each segment will be spread across the canvas, giving a result closer to simple histogram analysis.

In a preferred approach, good performance is achieved by setting the relative scaling factors of the co-ordinate axes so as to equalize the variances of the pixels in all the dimensions of the universe. An exception to this rule is that the canvas co-ordinates are arranged to have equal scaling determined by the wider (usually horizontal) co-ordinate.

Other embodiments employ methods of dynamically varying the relative scaling in order to minimize some error measure. For example, the coordinate axes may be scaled so as to minimize the product of errors evaluated along each axis, with the constraint that the scaling factors sum to a constant value.

In a preferred form of the invention, a distance measure is employed which takes into account the covariance of the image data.

The Euclidean distance measure used above can also be expressed in vector notation as $(\mathbf{x}-\mu).(\mathbf{x}-\mu)^T$. The so-called Mahalanobis distance measure can then be defined as equal to $(\mathbf{x}-\mu).\Lambda^{-1}.(\mathbf{x}-\mu)^T$ where Λ is the covariance matrix of the data in the segment, given by

$$\Lambda_{ij} = \frac{1}{K_S - 1}\sum_{k \in S}(x_{ik} - \mu_i)(x_{jk} - \mu_j)$$

Other variations on the basic Euclidean distance are also possible. One possibility, which is simpler to implement in hardware, is the Manhattan distance or L-1 norm. This is the distance between two points measured by walking parallel to the co-ordinate axes. Another possibility is to take the maximum of the differences measured along the co-ordinate axes.

In certain applications it will be useful to employ distance measures which take into account the covariance matrix, in ways other than the precise Mahalanobis distance, and which may additionally include components based on the the mass of the segment. In certain applications it will be advantageous to use a version of the Mahalanobis distance which is based on the covariance matrix of all the data, not taken segment by segment.

A simpler version of the Mahalanobis distance ignores the off-diagonal elements of the covariance matrix and simply compensates for the different variances along the coordinate axes.

The number of segments may be a parameter chosen by the user. In alternatives, the number of segments is chosen as a function of the input data. For example, the number of segments may be chosen so that the variance of the overall segmentation error approaches a predetermined constant value.

It is possible that a segment may disappear after some frames if its centroid turns out to be further from every pixel than some other segment. In many cases, this is desirable behaviour, as when an object disappears from the screen. In other embodiments, the mechanism may allow for the introduction of

new segments. Particular algorithms may establish criteria for splitting an existing segment into two, and possibly also for merging segments that end up being close together in the universe. Another approach to deciding whether to add or remove segments is to run two or more parallel versions of the algorithm with different numbers of segments and to base the decision on the difference in overall error between the two versions.

One solution to the problem of the disappearance and reappearance of objects because of global motion in the scene is to impose a toroidal structure on the canvas. This is done by stitching the left edge to the right edge and the top edge to the bottom edge. Centroids that disappear off one edge will now reappear at the opposite edge and will be available for re-use. Care needs to be taken in the distance definition to make sure that the shortest distance is used.

Certain embodiments use a pixel space in which the co-ordinates are simply the luminance, YUV or RGB values of the pixels. Others look at segmentation on the basis of motion vectors preferably derived in a phase correlation process. There are many other possibilities for features that can be included in the pixel space, for example local measures of noise or texture, or Gabor jets. For this reason, the pixel space (or the whole universe) is sometimes referred to as the feature space.

The description so far assumes that each pixel belongs to just one segment, so that assignment of pixels to segments is a hard decision. It is possible to replace this with a soft decision, in which each pixel carries a set of probabilities of membership of each segment. The output of the algorithm can be based on a hardened version of the decision (assigning the segment with the highest probability) while the soft decision is retained for the recursive processing. Alternatively, the soft decision may have significance in the output of the algorithm. For example, if the pixel space consists of motion vectors, the output may consist of several motion vectors assigned to each pixel, each with a relative weight.

In more detail, the action of choosing a segment for a pixel can be thought of as allocating a weight to the pixel's membership of each segment. In "hard" segmentation decisions, one of the weights is set to 1 and the others are set to 0, the choice being made according to which segment is "nearest" (by Mahalanobis, Euclidean or other distance measure) to the pixel concerned. A generalization of this is to make "soft" segmentation decisions, where the weights are constrained only to be between 0 and 1 and to sum to 1 across the segments. The weight for each segment could be calculated as a function of the distance measures. An example of a suitable function for pixel x is

$$p_m(\mathbf{x}) = \frac{e^{-d_m(\mathbf{x})}}{\sum_{m'=1}^{M} e^{-d_{m'}(\mathbf{x})}}$$

where M is the number of segments and $d_m(\mathbf{x})$ is the distance measure for segment m. If the overall model is based on multivariate Gaussian distributions, the weight calculated by the above function can loosely be interpreted as a "probability of membership" of each segment.

The weights can be used in various ways.

For example, in updating the segment centroids with new picture data, the centroid for each segment can be calculated as a weighted average of picture data across the whole picture, rather than an average across the segment. In this way each pixel contributes to a number of segments, each time in accordance with the probability that the pixel belongs to that segment. At the output of the process, a hard decision could be made for each pixel by simply choosing the segment with the maximum weight, but the weight could be retained for use as a "confidence measure" in further processing. For example, a motion vector output could be given a confidence measure which could be used to control a mix between a motion compensated signal and some fallback (e.g. non-motion-compensated) signal.

Alternatively, two or more (up to M) output values for each pixel could be retained together with their weights. For example, several motion vectors could be retained and each used to calculated motion compensated signal which could ultimately be combined by forming a weighted average.

Motion vectors give rise to the possibility of an additional component in the distance metric, based on the error in the pixel domain when the motion vector is applied to a picture. This displaced frame difference can be incorporated into the distance metric with appropriate scaling. The result is a segmentation that takes into account the fidelity of the motion compensated prediction.

Embodiments of this invention provide a number of important advantages. The described methods perform fast and (subjectively) very well on demanding picture material. The variation in the segmentation from one frame to the next is generally small; this is especially important for video. Moreover, segmentation decisions can frequently be taken in a single step. This not only reduces processing time but also increases reliability. With the scaling chosen, the segments are not necessarily contiguous but the parts of each segment remain close together. Replacement of "raw" motion vectors by a segmented version can improve the overall quality of the motion vector field.

It will be appreciated by those skilled in the art that the invention has been described by way of example only, and that a wide variety of alternative approaches may be adopted.

CLAIMS

1. A method of segmenting image data having a plurality of feature values at each pixel, in which the data is represented as points in a segmentation vector space which is the product of the vector space of feature values and the vector space of pixel addresses, wherein segments are represented as locations in the segmentation vector space, and the membership of a segment for each pixel is determined by the distance in segmentation vector space from the data point representing the pixel to the location of the segment.

2. A method according to Claim 1, in which the segments are represented as points.

3. A method according to Claim 1, in which the segments are represented as linear functions mapping the vector space of pixel locations to the vector space of pixel values.

4. A method according to Claim 1, in which the distance measure is a Euclidean distance.

5. A method according to Claim 1, in which the distance measure is a Manhattan distance.

6. A method according to Claim 1, in which the coordinate axes are scaled to equalize the variances of the data along each axis.

7. A method according to Claim 1, in which the coordinate axes are scaled in order to minimize the product of errors evaluated along each axis, with the constraint that the scaling factors sum to a constant value.

8. A method according to Claim 1, in which the distance measure is a Mahalanobis distance.

9. A method of segmenting image data having a plurality of feature values at each pixel, comprising the steps of representing the data as points in a segmentation vector space which is the product of the vector space of feature values and the vector space of pixel addresses; representing segments as locations in the segmentation vector space; determining a covariance matrix of the image data in each segment; measuring a distance in segmentation vector space of each pixel to each segment location taking into consideration said covariance matrix and determining the membership of a segment for each pixel through said distance measure.

10. A method according to Claim 9, where the covariance matrix Λ of the data in the segment is given by

$$\Lambda_{ij} = \frac{1}{K_S - 1} \sum_{k \in S} (x_{ik} - \mu_i)(x_{jk} - \mu_j)$$

where $(x_1, x_2, \cdots, x_N)_k, k \in S$ are vectors in the multidimensional space belonging to segment S, and the location of the segment is given by

$$(\mu_1, \mu_2, \cdots, \mu_N) = \frac{1}{K_S} \sum_{k \in S} (x_1, x_2, \cdots, x_N)_k$$

where K_S is the the number of points in segment S.

11. A method according to Claim 10, wherein the distance measure is equal to $(\mathbf{x} - \mu).\Lambda^{-1}.(\mathbf{x} - \mu)^T$.

12. A method of segmenting video data having a plurality of feature values at each pixel in a sequence of pictures, in which the data is represented as points in a segmentation vector space which is the product of the vector space of feature values and the vector space of pixel addresses, and wherein segments are represented as locations in the segmentation vector space, the method comprising the steps for each picture of initially assigning pixels to segments according to the segment membership of the respective pixel in the preceding picture in the sequence; calculating the location in segmentation vector space for each initial segment utilising feature values from the current picture and determining the membership of a segment for each pixel according to the distance in segmentation vector space from the data point representing the pixel to the location of the segment.

13. A method according to Claim 12, in which the segments are represented as points.

14. A method according to Claim 12, in which the segments are represented as linear functions mapping the vector space of pixel locations to the vector space of pixel values.

15. A method according to Claim 12, in which the distance measure is a Euclidean distance.

16. A method according to Claim 12, in which the distance measure is a Manhattan distance.

17. A method according to Claim 12, in which the coordinate axes are scaled to equalize the variances of the data along each axis.

18. A method according to Claim 12, in which the coordinate axes are scaled in order to minimize the product of errors evaluated along each axis, with the constraint that the scaling factors sum to a constant value.

19. A method according to Claim 12, in which the distance measure is a Mahalanobis distance.

20. A method of segmenting image data having a plurality of feature values at each pixel, in which the data is represented as points in a segmentation vector space which is the product of the vector space of feature values and the vector space of pixel addresses, comprising the steps of scaling the image data so as substantially to equalize the variance of the data in at least one dimension of the pixel address and each dimension of the feature value; initially assigning pixels to segments; representing each segment as a location in the segmentation vector space; and determining the membership of a segment for each pixel according to the distance the segmentation vector space from the data point representing the pixel to the location of the segment.

21. A method of segmenting image data having a plurality of feature values at each pixel, comprising the steps of representing the image data as points in a segmentation vector space which is the product of the vector space of feature values and the vector space of pixel addresses in a toroidal canvas; initially assigning pixels to segments represented as locations in the segmentation vector space, and determining the membership of a segment for each pixel according to a distance measure from the data point representing the pixel to the representation of the segment.

22. A method according to any one of the preceding claims, in which the feature values include pixel values and motion vector values.

23. A method according to Claim 22, in which the feature values include displaced frame differences.

24. A method according to any one of the preceding claims, in which each pixel is chosen to be a member of a single segment determined by minimizing the distance measure.

25. A method according to any one of the preceding claims, in which the number of segments is chosen by the user.

26. A method according to any one of the preceding claims, in which the number of segments is chosen as a function of the input data.

27. A method according to any one of the preceding claims, in which the number of segments is chosen so that the variance of an overall error measure approaches a predetermined value.

28. A method according to any one of the preceding claims, in which two or more parallel versions of the algorithm are run with different numbers of segments and the number of segments chosen is based on the relative performance of the two versions.

29. A method according to any one of the preceding claims, in which the representations of segments in the vector space are updated according to the segment membership of pixels.

30. A method according to any one of the preceding claims, in which the processes of assigning pixels to segments and of updating the representations of segments are repeated alternately.

31. A method according to any one of the preceding claims, in which the initial segmentation is taken from the previous picture in a sequence of pictures.

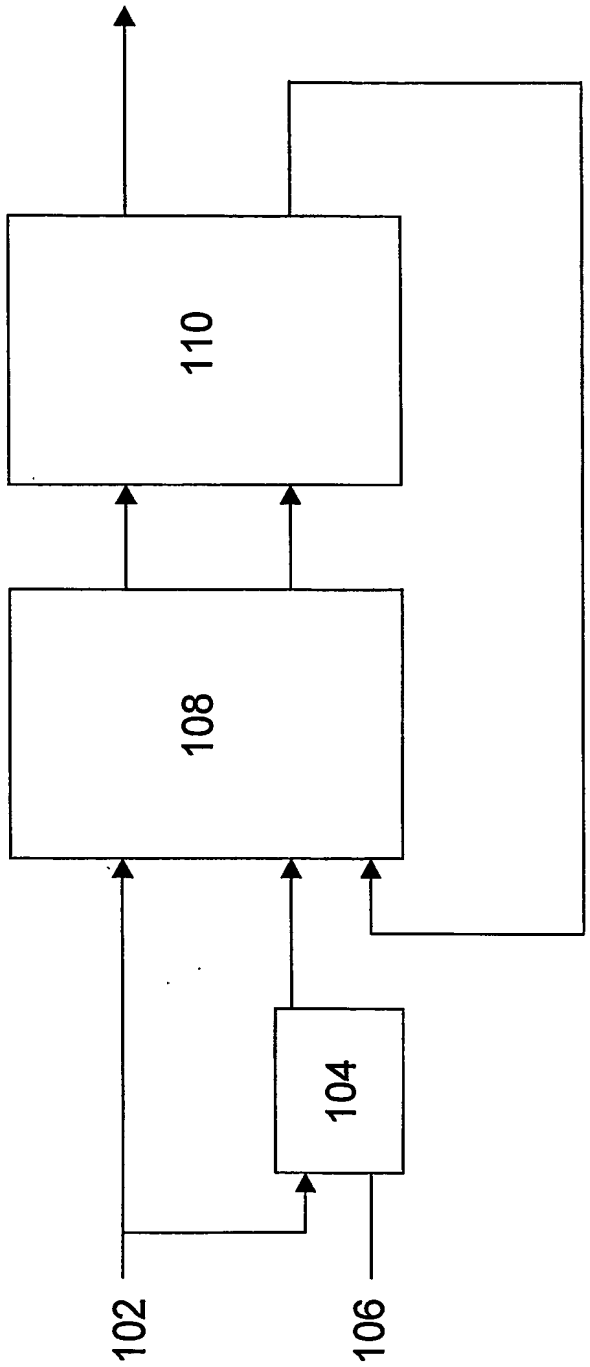


Fig. 1